

DOCUMENT RESUME

ED 163 605

EA 011 083

TITLE The Western Australian Surveys of Primary School Achievement, 1955-1970.

INSTITUTION Western Australian Education Dept., Perth. Research Branch.

PUB DATE [76]

NOTE 55p.

EDRS PRICE MF-\$0.83 Plus Postage. HC Not Available from EDRS.

DESCRIPTORS Achievement Rating; *Achievement Tests; *Comparative Analysis; Educational Research; Elementary Education; Foreign Countries; National Norms; *National Surveys; Norm Referenced Tests; Tables (Data); Test Interpretation; *Test Results; Test Validity

IDENTIFIERS *Australia

ABSTRACT

This report reviews the results of the Australian Surveys of Primary School Achievement given every five years from 1955 to 1970. It compares results of the tests and explains why the survey was discontinued in 1975. Subjects were approximately 1,000 fifth-grade and 1,000 seventh-grade students selected as a representative sample of Australian school children. Comparison of the tests reveals 1960 to be the peak performance year, with the 1970 test results being somewhat higher than 1965 results in some areas. Overall scores for all years are higher for girls, for classes of 40 or more students, for metropolitan rather than country schools, and for single-grade rather than multigrade classes. Research problems that qualify results include curriculum changes over the years, different aims of different tests, and the fact that tests were not administered at the same time during the school year. In 1975 the surveys were discontinued and replaced by another test because the test battery was outmoded, the test failed to discriminate among the majority of students, survey aims were too ambitious, and because researchers favored a criterion-referenced approach over the norm-referenced approach used in the surveys. (JH)

* Reproductions supplied by EDRS are the best that can be made *

* from the original document. *

ED163605

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

THE WESTERN AUSTRALIAN
SURVEYS OF PRIMARY SCHOOL
ACHIEVEMENT
1955-1970



"PERMISSION TO REPRODUCE THIS
MATERIAL IN MICROFICHE ONLY
HAS BEEN GRANTED BY

M. J. Angus

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC) AND
USERS OF THE ERIC SYSTEM."

RESEARCH BRANCH
EDUCATION DEPARTMENT OF WESTERN
AUSTRALIA

TABLE OF CONTENTS

	<u>Page</u>
PREFACE	v.
AIMS	1
The Earlier Surveys	1
The Concepts of Retardation and Backwardness	2
Aims of the 1970 Survey	4
SAMPLING PROCEDURES	5
The Importance of Sampling	5
Description of Procedures Employed	5
Discussion of the Sampling Procedures	6
The Omission of Junior High Schools From the 1970 Sample	8
Missing Scores	9
Characteristics of the 1970 Grade 7 Sample	9
THE TEST BATTERY	11
Reliability	11
Validity	12
Speededness of the Tests	14
Discrimination of the Tests	15
Intercorrelations Among the Tests	16
RESEARCH STRATEGY	17
Methodological Limitations	17
The Validity of Cross Year Group Comparisons	17
RESULTS	19
Year Group Comparisons: Grade 7	19
Year Group Comparisons: Grade 5	23
A Multiple Discriminant Analysis of the Grade 7 Results	24
Grade 5 and Grade 7 Comparisons	27
The 1960 Peak	28
The Intelligence Test Results	29
The Relationship Between School and Other Factors to School Achievement	30
CONCLUSIONS	
The 1955 Survey	35
The 1960 Survey	35
The 1965 Survey	36
The 1970 Survey	37
REFERENCES	40
APPENDICES	41

PREFACE

The Context of this Report

Nobody seriously disputes that whatever else schools accomplish they should at least produce graduates who are functionally literate and numerate. Yet, the teaching profession and the public have always been aware, not always happily, of students who according to some usually ill-defined criterion have not acquired sufficient facility in these basic skills. The issues that seem seasonally to re-emerge pertain to the *discovery of evidence* relating to standards of school performance, the *extent* of numeracy and literacy in schools and tertiary education institutions, and to the *causes* of this phenomenon. The media seem always keen to highlight reports of a decline in standards and attribute with utter conviction reasons for the decline to various convenient explanations.

Establishing clear cut trends in reading or arithmetic standards by means of large scale surveys has not been a successful enterprise over recent years. Much of the problem lies in the failure of researchers and the public to arrive at a definition of what constitutes a satisfactory standard of performance in literacy and numeracy, or even the more specific fields of reading, writing and arithmetic. In addition the research methodology inevitably appears flawed when received in calm of hindsight. Burke and Lewis (1975) reviewed six major studies of the reading standards of British school children and concluded that "No justification was found for a belief that standards are declining" (p.163). A number of these studies have been widely cited as evidence that numeracy and literacy are on the wain.

In this context it is not surprising then that the *Bullock Report* adopted such an equivocal stand when analysing existing evidence. In its summary, after examining available British Research evidence relating to standards of literacy, the Report concluded in the following terms:

There is no evidence of a decline in attainment over the years in the lowest achievers among 15 year olds. (p.517)

The Report also concludes that:

There was no significant change in the reading standards of 11 year olds over the decade 1960-1970, but such movement as took place after 1964 was in all probability downwards. (p.517)

Furthermore, it notes:

There is some evidence that children of seven are not as advanced as formerly in those aspects of reading ability which are measured by tests. (p.517)

It is not surprising that when such a prestigious group of professionals as the Bullock Committee must report such qualified findings that the public looks towards other more speculative evidence.

This is the dilemma in which professional researchers so often find themselves. When the data does not fall into a consistent, predictable and clearly identifiable pattern (and it very rarely does) qualifications and reservations are required. For the researcher to ignore any weakness in the survey design or blithely discount a finding that is at odds with the general conclusion would be an act of irresponsibility.

If judgments must be made regarding decline or improvement in basic skills over a period of time then it is critical that accurate and meaningful base line data is acquired. In 1975, the projected year for the fifth Survey of Primary School Achievement a national study was initiated which appeared to have eliminated many of the flaws that had limited the Education Department surveys. At the request of a Select Committee of the House of Representatives, the Australian Council for Educational Research initiated the Australian Study of School Performance, a survey of the basic learning skills of Australian school children. Over 6,000 ten year olds and 6,000 fourteen year olds were sampled across all States and Territories. The tests were oriented towards measuring performance on what could be called fundamental "survival" skills; completing a form, understanding a newspaper advertisement, calculating profit from a simple business transaction, and the like. While data are not available from this study which permit inferences to be made on the relative standing of the 1975 ten and fourteen year age cohorts with their 1970 or 1965 counterparts, cross national comparisons suggest that the Australian students perform at comparable levels to their British and American age peers. The study does however offer a useful beginning point of researchers concerned with changes of standards over the next decade.

The Surveys of Primary School Achievement

The 1955 Survey of Primary School Achievement was, as were the subsequent Surveys, administered at Departmental initiative for the overall purpose of monitoring achievement in various language and arithmetic skills. The 1960 survey results were published in the Education Circular of the following year. The reports of the 1955, 1965 and 1970 surveys, while available on request, were not disseminated widely outside the Department.

The 1970 survey report is being published at this point in time for a variety of reasons. As the euphoria that surrounded the educational initiatives of the first half of the 1970's gradually dissipates there are renewed calls for evidence relating to standards of work in schools. While the 1970 survey was addressed to standards in schools prior to the present decade, never-the-less it can contribute to the debate as readers will see for themselves.

In addition, the present report demonstrates a significant attempt by the Education Department to review by objective means as possible the level of standards in its schools. Various voices have asked what the Education Department has done to discover the situation for itself. The 1970 report as well as detailing the basis of the four statewide surveys also explains why the survey was discontinued in 1975.

Notwithstanding the present public interest in the question of changes in standards the decision not to replicate the survey for a fifth time in 1975 appears to have been a sensible one. The survey test battery is clearly outmoded. Notions of criterion referenced testing, which were incorporated in the Australian Council for Educational Research, Australian Study of School Performance to which reference has been made, were relatively novel ideas in this country in 1970 and virtually unheard of in 1955. The tests which comprised the initial 1955 survey test battery and which were reused in subsequent surveys were a "mixed bag" of diagnostic and norm-references instruments. For example, the arithmetic tests were devised in order to detect the sources of error in the four computational processes of adding, subtracting, multiplying and dividing; or in the case of the reading test, to compare students' performance with that of their age or grade peers. The tests were not constructed with a view to establishing the numbers of children who could complete the essential skills or processes.

It is easy in hindsight to perceive weaknesses in survey strategy or data analysis techniques beyond those acknowledged in the report. For example, in a number of instances statistical tests of differences were not conducted where they might have been conducted. However, it must be remembered that the growth in sophistication of computer techniques over the past five years has been remarkable. In 1970 the P.D.P. 6 computer at the University of Western Australia, by which means the data were analysed, simply did not have the capacity to process the volume of data according to the multivariate statistical techniques which were deemed appropriate. Yet in 1975 such data processing on the larger computer which has subsequently been installed would have appeared a relatively straight forward exercise.

It has been suggested that more could have been made of the survey data than was actually described in the 1970 Report. For example, a further analysis of the "class size" phenomenon whereby at the Grade 7 level children in larger classes outperformed their peers in smaller classes may now appear warranted. One might have cared to cross-tabulate the "class size" variable with the "class of school" variable to see whether the trend still favoured larger classes in all sizes of schools. No doubt readers will devise and propose a new range of possible analyses.

Accepting the limitations of the tests employed and the restrictions imposed by the sampling strategy, the Survey of Primary School Achievement should be regarded as important evidence relating to standards in a limited range of language skills. The surveys represent the only objective attempt to monitor standards from 1955 to 1970 at the Grade 7 level in Government primary schools in this State. The results cannot be interpreted unequivocally. The 1960 results are sufficiently anomalous to induce a note of caution in any generalisation. The slight improvement in 1970 in reading comprehension, over the 1955 and 1965 results precludes any overall conclusion regarding decline in standards. The reading comprehension test was one of the more adequate tests in the battery.

One further motive in releasing the 1970 survey report is to encourage among its readers an attitude of caution in appraising allegations of declining or improving standards. It should be appreciated that to establish clearly the situation requires an investment of research skills, patience

and resources that are rarely available for that purpose. It remains to be seen whether advances in research methods and the demands of the public teaching profession will produce means by which more compelling evidence can be assembled.

M.J. ANGUS
RESEARCH BRANCH
May 1976

Bourke, Elizabeth and Lewis, D.G. Standards of Reading: A Critical Review of Some Recent Studies. *Educational Research*, 1975, 17(3), 163-173.

A Language for Life, Department of Education and Science. Report of the Committee of Inquiry appointed by the Secretary of State for Education and Science under the Chairmanship of Sir Alan Bullock, F.B.A., B.A., London, H.M.S.O., 1975.

Survey of Acceleration and Retardation, 1960. *Education Circular*. Education Department of Western Australia, 1961, 63, 222-224.

AIMS

The Earlier Surveys

The Survey of Primary School Achievement or Survey of Retardation and Acceleration as the project has been known, was first carried out in 1955. The statement of aims of the project in that year was as follows:

- i. Conduct a survey of research findings on retardation and acceleration.
- ii. Conduct a survey of the extent of individual differences of Western Australian children and of the extent of acceleration and retardation of children's attainments in the skill subjects as against -
 - (a) intellectual ability.
 - (b) grade.
- iii. Research into the best ways of dealing both in school and class organisation and in teaching techniques with the children not working up to their ability - both bright and not so bright and the children whose retardation is such that sound methods by the schools within its resources should be reasonably expected to apply diagnostic and remedial work.
- iv. Research work into the causes of retardation in the primary school.
- v. Research into the best methods of dealing with children retarded more than two years both as to methods and as to organisation.

The authors of the 1955 survey report recognised that the research brief as outlined above could not be effected unless a large scale project was implemented on a longitudinal basis, and that as a consequence, only stages (i) and (ii) could effectively be incorporated into the research project for that year. Plans were made to continue the survey of performance at Grade 7 level at five yearly intervals.

Five years later "catering for individual differences" was the prominent theme of the 1960 survey of retardation and acceleration. The brief of this project was almost identical to that of the 1955 survey except that attention was now drawn to the relationship between size of class and type of class, and acceleration and retardation.

More particularly, the 1960 survey sought to discover the degree of retardation and acceleration in arithmetic and reading of children in Grades 5 and 7 throughout the State with a view to:

- i. comparing the results of the present survey with a similar survey carried out in 1955 of acceleration and retardation in arithmetic and reading of Grade 7 children;
- ii. ascertaining possible factors which have led either to improvement or deterioration in standards of attainment;
- iii. discovering which areas of attainment still require considerable attention with regard to raising standards; finding out what possible effects size of class and type of class have on standards of educational achievement.

The 1965 survey differed from the earlier surveys not only in title (it was identified as The 1965 Survey of Primary School Achievement) but also in emphasis. The aims of the survey were to assess the achievement of Grade 7 primary school pupils in selected subjects with a view to:

- i. determining the influence of selected factors such as location and size of school, school organisation, class size, bus travel and sex of pupils;
- ii. comparing the results with those obtained in the surveys conducted in 1955 and 1960 in the norming of the tests in 1946; and
- iii. establishing new norms for certain Australian Council for Educational Research tests commonly used in schools.

Briefly, the 1965 survey in large measure discarded the notions of backwardness and retardation and sought to look at the relationships between achievement and certain school factors. This decision had important implications for the 1970 survey. In accomplishing its third purpose, the norming of the Australian Council for Educational Research tests a decision was made to include a sample of Grade 5 children in addition to the Grade 7 sample.

The Concepts of Retardation and Backwardness

Retardation was defined as not working up to *intellectual ability*. Acceleration occurred when results in the skill subjects exceeded what could be expected from the intelligence test results. *Backwardness* was defined as not working up to the *grade level* even though achievement may have equalled intellectual ability.

The concepts of backwards and retardation received considerable publicity and impetus after the publication, in 1942, of Schonell's book *Backwardness in the Basic Subjects*. However in the late 1950's and during the 1960's criticism of the notion of backwardness appeared in the professional literature. For example Vernon (1956, p.1277) writes:

The inescapable conclusion is that the whole conception of retardation relative to intelligence is so unstable, so dependent on the particular test employed, as to be almost meaningless psychologically and educationally.

Crane (1959, p.258) points out that it would be absurd to deny that educationally retarded children do not exist. Rather, he questions

...the validity of much current practice in the detection of those who are retarded and the massive confusion of thought in which the concept of retardation has become inextricably snared.

Thus, although retardation and backwardness seemed useful concepts for the purposes of the 1955 and 1960 survey by 1965 their value seemed questionable. Apart from seeming simplistic in the light of advances in learning theory the application of these concepts in large-scale surveys involved complications of a statistical and measurement kind. For example, assessing the degree of retardation and backwardness requires the conversion of test scores to age or grade norms. Such norms make assumptions about learning rates that are difficult to verify (see Flanagan (1951) for a discussion of this issue) and which are not necessary for the purposes of monitoring achievement. Furthermore, employing the concepts of retardation and backwardness raises the problem of statistical regression. Whenever two variables (such as performance on reading and intelligence tests) are not perfectly correlated then students who score highly (or poorly) in one test are more likely to score closer to the mean performance on the second variable or test. In other words the higher performers on an intelligence test are unlikely to receive the highest score on a reading test. This apparent underachievement (or retardation) on the part of these students is the result of the statistical artefact, regression towards the mean. Hence evidence of *underachievement* of individual children, gathered from the single administration of tests may not always be valid.

Aims of the 1970 Survey

The general aims of the 1970 survey were two-fold:

- i. to look at patterns of student performance in the basic skill areas of reading, spelling and arithmetic, and
- ii. to investigate the influence of the factors school location, sex, class of school, age, class size and class grouping on student achievement in reading, spelling and arithmetic skills.

The original data from 1955 and 1965 were available for re-analysis. The summaries of mean or median performance for the three previous surveys, 1955, 1960 and 1965 were also available. Hence there exists the opportunity to review standards of performance of Grade 7 pupils from 1955 to 1970 and of Grade 5 pupils from 1965 to 1970. As well, other characteristics of the biography, home background, and teaching procedures for each individual child were available. It was the intention to associate these characteristics with the performance of the children on the tests of basic skills and to thereby determine their relative influence on the learning of the basic skills.

SAMPLING PROCEDURES

The Importance of Sampling

If investigators wish to discover deterioration or improvement in basic skills at a particular grade level then, assuming that valid testing arrangements can be effected, a relatively conclusive research approach would be to measure the performance of every Grade 7 pupil. However, such an exercise would be extremely costly and disruptive. Instead one means or other of sampling is usually employed. The efficacy of a sampling procedure can be estimated, cost factors aside, by the extent to which the sample statistics acquired truly represent the population parameters sought. If there is a wide discrepancy, whatever the cause, and if the extent of the discrepancy is unknown, then the whole sampling and survey exercise will most likely have been wasteful or even harmful. However, estimating the efficiency of the sampling procedure adopted is a difficult process in itself. More often than not a good deal of guess work is employed in ensuring that sufficient precision of sampling has been effected.

The intention of the survey was to provide evidence that would allow readers to infer generalisations from the samples to the entire State-wide populations of Grade 5 and Grade 7 children. The importance of adopting sound sampling procedures for the Survey of Primary School Achievement cannot be over estimated. There are so many possible correlates of successful school performance that the omission or over emphasis of a segment of the population that is separately influenced by one of these correlates will provide scope for a biased sample and consequently, biased conclusions.

Description of Procedures Employed

In general, comparable sampling strategies were adopted for each of the 1955, 1960, 1965 and 1970 survey. The procedures are detailed below:

- i. The total enrolments for each class of school was found and presented as a percentage of the total enrolment.
- ii. These percentages were then applied to the sample of approximately 1,000 in each grade to ascertain the number to be listed in each class of school.

6.

- iii. To obtain the proportion of metropolitan and rural children to be tested, the number of schools in each class of school was separated on percentage basis.
- iv. The percentages mentioned in (iii) were then applied to the number of each class of school to be tested as obtained in (ii).
- v. The process was repeated (unlike the 1955, 1960 and 1965 procedure) separately for both the Grade 7 and Grade 5 samples, that is a Grade 7 and a Grade 5 sample of schools was drawn up.
- vi. A sufficient number of schools was then chosen from each strata that would meet the required proportions of students as detailed in Appendix 1. This number was fixed at about 1,000 students at each grade level. The number of 1,000 was chosen somewhat arbitrarily on the basis of earlier survey sample sizes and with the hope that this sample size (approximately seven per cent of the State population) would be representative of the population as a whole.

Discussion of the Sampling Procedures

Thus, the sampling procedures were characterised by major departures from simple randomised sampling of pupils.

- i. Fixed proportions of students were assigned to certain pre-determined strata, namely "class of school" and "location of school", according to enrolment proportions.
- ii. The sampling unit employed (that is the element that was actually randomised) was *school* not pupil.

Statistics are used to make inferences about the population. The more adequate or representative the sample the more accurate these influences are likely to be. If we want to have some previously specified confidence in our inferences then the sampling ought not be done on a haphazard, "take-what-comes" basis. Random sampling where each element has an equal chance of being selected, is commonly employed and if carried out according to all the rules, should meet the requirements for drawing inferences about population parameters. However it should be noted that there will probably be sampling

error even in random samples, and thus, even a rigorously drawn random sample could provide substantially erroneous estimates of the relevant population parameters.

A strict, random sample of Grade 7 or Grade 5 children would require the listing of every Grade 7 or Grade 5 child in the State and arrangements made so that every individual child had an equal opportunity of selection. Such a procedure, though desirable was considered impracticable. Instead, the sampling unit was the school Grade 5 or Grade 7 cohort of pupils. Each unit or class of pupils was chosen randomly from a register of schools. This form of sampling is known as cluster sampling. It is a less efficient means of obtaining a representative sample than simple random sampling though it has obvious practical advantages. Usually, to improve the efficiency of cluster sampling a larger number of subjects are selected than were simple random sampling employed. Ideally, the larger the number of units (schools or classes) selected the better.

For each of the four surveys stratified cluster sampling of schools occurred. At least two variables determined the strata: class of school (that is, a grading of the school based largely upon its size) and location of school (a categorisation of school location according to its location in either the country or metropolitan area. An assumption of the four surveys was that these two factors, location and class of school, would explain differences in standards of pupil performance between schools. Ordinarily, one would expect a proportionate, stratified sample to provide a more precise or representative estimate of the population values than a simple random sample would provide. The gain in precision (as compared with a simple random sample) is related to the correlation between the variable being studied (in this case pupil performance) and the variable used in the stratification.

Sex of students was used as a fourth stratum in the 1965 sample. Thus slightly different sampling procedures were adopted in 1965 for earlier surveys and the 1970 survey. It was thought that the gain in precision that might be won from employing the additional two strata would be minimal. The decision in sampling as to whether to employ more than one or two strata is analogous to the decision that arises in applications of multiple prediction in which one almost always obtains rapidly diminishing returns as one increased the number of predictors. This is

due to the fact that the various strata factors are themselves correlated and each additional variable after the first one or two predicts only a small amount of criterion variance not already explained.

The Omission of Junior High Schools From the 1970 Sample

One important discrepancy between the initial four survey samples was the omission from the 1970 sample of Junior High School Grade 5 and 7 pupils.* In order to effect comparisons between the 1970, 1965 and 1955 data the Junior High School data had to be omitted during the data analysis. Hence, results of all comparisons in this report cannot strictly be generalized to the Junior High School Grade 5 and 7 populations.

The 1965 Junior High School sub-sample was separately processed. The means and standard deviations of the Junior High School sub-sample are detailed in Table 1 showing corresponding statistics for the rest of the 1965 survey sample.

TABLE 1: MEANS AND STANDARD DEVIATIONS OF JUNIOR HIGH SCHOOL AND REMAINING 1965 GRADE 7 SAMPLE

	1965 Junior High School (N = 122)		1965 Total Sample - Excluding Junior High School (N = 1149)	
	$\bar{X}$	S.D.	$\bar{X}$	S.D.
I.Q.	101.31	11.68	102.21	12.61
Word Knowledge (10 min)	39.99	14.51	42.96	13.87
Word Knowledge (15 min)	49.66	16.06	54.74	16.99
Reading for Meaning	32.38	11.76	35.49	11.33
Addition	44.65	5.30	44.80	5.79
Subtraction	47.53	10.33	49.98	9.87
Multiplication	32.90	4.95	33.22	4.86
Division	32.95	8.49	33.39	8.31
Spelling (words)	41.13	9.51	43.27	8.14
Spelling (sentences)	25.04	10.00	28.14	11.00

* Junior High Schools are small rural high schools which have a primary school and secondary school component catering for three years of post primary education. They are usually located in those communities which do not have sufficient number of post primary pupils to warrant a high school with facilities for five years of post-primary education.

The results show consistent small differences which do not favour the Junior High School sample. These results tend to justify the removal of this section of the sample from analyses which compare the performance of pupils from all four surveys.

Missing Scores

A third factor which had implications for the sampling procedure and subsequent data analysis was the omission from the data analysis of incomplete student results. It was felt that the complications that would accrue from making special provision to process student results with missing test scores were formidable. There appeared to be no systematic bias influencing students who had incomplete results. Pupil absenteeism was responsible for most incomplete results. The practice of discarding incomplete results was followed in all four surveys. Between ten and 15 per cent of the initial sample were not further considered for analysis.

Characteristics of the 1970 Grade 7 Sample

Grade 7 Sample

The 1970 sample was stratified on only two factors, school location and type of school. As the purpose of the survey was to investigate differences between factors such as school location, organisation, class size, it is interesting to note the proportions of these sub-classes which were included in the final sample. The respective proportions for the 1965 sample are also detailed for consideration in Table 2. The 1965 and 1970 sample structures need not necessarily be identical as administrative differences in school organisation would have an impact in this regard.

Particularly large differences are noticeable in the streamed/unstreamed and class size categories. In both instances the difference reflects a change in attitude towards school organisation. In the latter instance, in the 'population' of primary schools 42.3 per cent of classes contained 40 students or more in 1965 and by 1970 this percentage had fallen to 22.1 per cent.

TABLE 2: BREAKDOWN BY PERCENTAGE OF THE 1965 AND 1970 GRADE 7 SAMPLE
(N = 1149 and 1108 respectively)

		1965	1970
Sex	Boys	49.1	48.5
	Girls	50.9	51.5
Location	Metropolitan	68.1	55.4
	Country	31.9	44.6
School Organisation	Mixed Grades	22.4	20.7
	Unmixed Grades	77.6	79.3
	Unstreamed	37.0	63.1
	Streamed	63.0	36.9
Class Size	0 - 20	2.4	1.7
	21 - 30	10.9	10.9
	31 - 40	23.8	47.6
	41 +	62.8	39.7
School Type	IA	25.3	38.8
	I	49.3	40.0
	II	12.4	8.6
	III	11.8	10.2
	IV	1.2	2.4

A comparison of percentages of the sample broken down into classification of school for the 1960, 1965 and 1970 surveys is shown in Appendix 1.

In addition this appendix contains the parameters for Government primary schools.

THE TEST BATTERY*

Reliability

With any testing programme of this sort the suitability of the test instruments is of paramount importance. If the tests fail to measure what they are *purported* to measure or what they are *required* to measure (that is, they are invalid), or if they measure with limited accuracy (that is, they are unreliable) then the survey would be a wasteful exercise. Usually the extent of reliability and validity is a matter of degree. One can only arbitrarily decide that a test is too invalid or too unreliable.

A test is reliable when it measures a student's abilities consistently. If a student completed the same test on two different occasions and fatigue, practice effects and marking variations aside, the student obtain different results, then this difference reflects the unreliability or inconsistency of the test. The test results of an unreliable test contain such large error components that any statement about the relative standing or achievement of persons who completed the test is fraught with hazards; statements about the student's achievement are about as valuable and accurate as statements about the student's height obtained with an elastic ruler.

A table of tests and reported (test manual) reliabilities, length, and time limits is contained in Appendix 2. Unfortunately no estimates of test reliability were conducted with the 1970 survey (or earlier surveys) as tests were hand marked in schools using stencils and only total student scores were returned to the Education Department. The reliability of the test is to some extent determined by the number of test items, the item difficulty, the speededness and the discrimination of the individual test items, and by the heterogeneity of the distribution of test scores within the group tested. Hence the reliability figures (reported as coefficients ranging theoretically from zero to one) shown in Appendix 2, which were abstracted from the test manuals, may not be accurate estimates since the estimates may vary from sample to sample.

* Appendix 2 contains a list of tests employed in each of the surveys. Appendix 3 contains illustrations of the type of items used in tests administered in the 1970 survey.

Reliability data is not available for locally produced tests used in the four surveys. Of the Departmentally constructed tests the only ones which are of interest in this Report, are the Western Australian Spelling Tests. These tests were composed by taking random samples of fifty words from the *My Word Books* for Grades 5 and 7.

One means of estimating the reliability of a test is to observe its correlation with a 'parallel' or identical test form which measures the same traits and behaviours. Data abstracted from the 1965 survey allows estimates of reliability to be made as in that survey parallel forms of the Australian Council for Educational Research Word Knowledge and Reading for Meaning were deliberately administered for norming purposes. The correlations between Forms C and D for Word Knowledge and Reading for Meaning were 0.78 and 0.76. These estimates are not as high as those values typically reported in manuals of tests of a similar type.

In general, however, it seems reasonable to conclude that the reliabilities of the various surveys tests referred to in this report are adequate, or at least as high as the reliabilities of tests employed in comparable surveys.

Validity

The determination of the validity of a test battery is a difficult task. The concurrent validity of the tests may be established by measuring the degree of association between the test and a measure of the same construct - usually another test. With achievement tests where this procedure is not possible, one can systematically analyse the content of the course that is to be tested and then see whether the key facts, understandings, skills, etc. are represented among the test items. If the purpose of a test programme is to test arithmetic skills, but of the many types of skills only a restricted few are sampled by the test items, then the test is to some extent invalid.

When the first survey was conducted in 1955 the Achievement Testing "Movement" had not really got underway. Diagnostic tests as popularized by Schonnell and others predominated among classroom testing. For example, as far as testing individual arithmetic skills in addition, subtraction, multiplication and divisions and highlighting the particular skills that

each of these processes require. However, most teachers of primary school mathematics would argue strongly that there is more to mathematics than acquiring skill in these important, though rudimentary processes. One has only to read the aims of the Primary School Mathematics Syllabus to see this.

TABLE 3: SELECTED AIMS FOR MATHEMATICS: TAKEN FROM THE PRIMARY SYLLABUS OUTLINES, EDUCATION DEPARTMENT OF WESTERN AUSTRALIA, 1968

<u>Senior Grades (6-7)</u> 1. Continue to develop favourable attitudes towards mathematics understanding and skills. 2. Maintain and extend interest in the arrangement of form in the environment. 3. Continue to develop an awareness of the place of mathematics in society. 4. Develop an appreciation of the importance of mathematics and its growth from and for human needs.
1. Develop an understanding of the structural relationships between spatial elements. 2. Continue to develop a knowledge of the practical application of standard units of measure. 3. Extend the knowledge of the whole numbers and rational numbers and begin to develop a knowledge of the negative numbers and their properties. 4. Extend the knowledge of the operations on the whole numbers and rational numbers and the properties of these operations. Begin to develop a knowledge of the operations of the negative numbers. 5. Consolidate and maintain a knowledge of the basic sums, products, differences and quotients which apply to the decimal system of numeration. 6. Extend the knowledge of the terms, language and symbols of mathematics.
1. Continue to develop a skill in computation of the whole numbers and rational numbers with accuracy and speed. Begin to develop a skill in computation using negative numbers. 2. Continue to develop a skill in the application of numbers and other acquired mathematics concepts. 3. Continue to develop the ability in practical quantitative situations to make sound estimates and approximations to verify these. 4. Maintain and extend the ability to work in simple inequalities as well as equalities. 5. Extend the ability to think and communicate effectively in quantitative and other mathematical situations. 6. Continue to develop the ability to frame and test simple mathematical hypotheses.

The point being made in Table 3 is that the Australian Council for Educational Research Arithmetic Test is a measure of four fundamental processes and is not a measure of mathematics skills generally. Thus, though the Surveys may make reference to improvements or deteriorations in performances on these tests, no light is shed on the achievement of the broader aims of the mathematics. Hence as measures of the broader aims the tests are quite invalid. However no claim is made in this report for the validity of the tests under these broader circumstances. By the same token, the spelling and reading tests are not fully representative of language skills or even of all the various skills necessary to successfully spell and read, and, are only partially valid. This limitation is of course true of most written achievement tests though the extent of the invalidity is a matter of degree.

Speededness of the Tests

Most achievement tests are to some extent speeded. A test is completely unspeeded when the students standing is not altered if given additional time. The decision regarding how speeded a test ought to be is to some extent an arbitrary one. If the intention of the test administrator to reward the student who can complete the test items in a limited and pre-specified time, and if the test items are of moderate and homogeneous difficulty then the administrator is likely to validly achieve his objectives. The survey tests have a large speededness component built into them owing to the large number of items to be completed in a limited testing period. The question arises as to whether the concern of the survey ought to be with how many items students can answer in a limited amount of time. Some commentators in 1971 believe that the emphasis in speed work is a less important objective in 1970 than it was in 1955. Hence the tests, particularly the arithmetic and word knowledge components of the battery, while perhaps more valid tests of 1955 and perhaps 1960 pupil achievement may be less so of 1965 and 1970 pupil achievement on the grounds of their speededness bias. Surprising, however, an empirical investigation into the degree of speededness in the word knowledge test indicated the speededness, per se, played very little part in determining the ordering of students, at least for 1965. This matter is further discussed in Appendix 4.

Discrimination of the Tests

An inspection of the frequency distribution of test responses revealed a small dispersion of scores for the arithmetic tests, particularly at the Grade 7 level. A section of the 1970 Grade 7 multiplication scores is shown in Table 4.

TABLE 4: EXTRACT FROM 1970 GRADE 7 MULTIPLICATION FREQUENCY DISTRIBUTION

Score	Frequency	
.	.	
.	.	
.	.	
26	32	
27	27	
28	41	
29	52	No.Items = 55
30	74	N = 1026
31	59	S.D. = 5.9
32	87	
33	100	
34	139	
35	109	
36	79	
37	41	
38	23	
.	.	
.	.	
.	.	

In fact, 50.1 per cent of the sample scored between 32 and 36, a range of five score points. It would seem that the inability of this test to discriminate among the majority of the students ought prejudice to some extent comparisons between year groups and other sub-samples in the study. Presumably the test items are too easy and even under the speeded conditions fail to discriminate. The top score on the test was 41. Unfortunately the students' individual test papers were not obtained

from schools so that an individual item analysis for each test was not possible. The item reliabilities may have provided relevant data regarding the efficacy of the test as a whole.

What is most important and which can be determined from Table 4 is that unfamiliarity with a single item could seriously alter the whole distribution and bias the performance, significantly of a particular survey year group. The spelling and reading tests would yield a case in point.

Intercorrelations Among the Tests

An examination of the correlation matrices for the 1965 and 1970 Grade 5 sample revealed that all tests were at least moderately correlated with each other, indicating that the battery as a whole may have been measuring a general achievement factor. In the 1965 sample the lowest test inter-correlation coefficient was 0.39 between *spelling* (words) and *addition* and the highest was 0.78 between *word knowledge* and *reading*. For the 1970 sample, the lowest inter-correlation was 0.31 between *I.Q.* and *addition* while the highest was 0.73 between *spelling* (sentences) and *word knowledge*.

RESEARCH STRATEGY

Methodological Limitations

From one point of view, the 1970 Survey of Primary School Achievement represents the fourth stage of a longitudinal study of Grade 7 achievement. Implicit in such a project are numbers of research problems. Most can be attributed to three main sources: the selection of fair and representative samples on each of the four occasions, the derivation of reliable and valid monitoring devices, and the control (or lack of control) of a host of factors that might explain variation in performance from survey to survey. As a consequence, the researcher is placed in a position where he must attempt to establish whether differences in the mean performance of each of the four samples are due to differences in teaching strategy, improved educational facilities, and so on, or are explained by fluctuations caused simply by measurement or sampling error.

Probably the aims of the survey are too ambitious. It would have been a sufficient accomplishment to accurately detect changes in standards of a prescribed set of skills. To determine the influence of various school factors introduces a host of complications and research issues. Studies of the influence of class size or streaming have been conducted over the past decades without resolving the basic questions. This is not to say that it should not be a legitimate area of interest for an administrator to discover, for example, whether children in country schools perform on tests with equal competence to their city counterparts. The point being made is that the survey is not the appropriate vehicle for discovering *why* this phenomenon occurs; nor does it necessarily provide leads to remediation.

This research strategy of the 1970 survey has been to *describe* events rather than attempts to *explain* patterns of school performance.

The Validity of Cross Year Group Comparisons

The report of the 1965 survey drew attention to three major limitations on comparisons of performance between year groups.

Firstly, the past twenty years have seen a changing school population. Chronological promotion has meant a drop in the average ages of pupils in various grades - those in 1965 being 1.5 months younger than those in 1955. It follows, therefore, that the samples used differ. The total populations from which they were selected were different and the sampling techniques used also differed.

Secondly, there is the question of the present validity of the norms of the A.C.E.R. tests. These were designed and originally normed in 1946, since which time, curricula have changed. Further, there has been a change of emphasis in the curriculum. This has meant less emphasis on the traditional skill subjects and more on social studies and art. When the tests were normed, the norms were established on the total population including non-Government schools. The results of the 1965 survey only concern children in Government schools. The Silent Reading Tests were used with separate answer sheets in 1965 while the original norming, answers were written directly into the booklets.

Thirdly, there is some variation in the time of testing. The original norming of the A.C.E.R. tests was carried out in November 1946. The 1955 survey was held in the first week of August, the 1960 survey in the last week in October and the 1965 survey in the first week of November.

While there is substance in these limitations they do not necessarily preclude useful cross survey comparisons providing there is some awareness of their potential influence. They represent only three of a long list of potential intervening influences that could explain differences in performances.

For example, one may speculate that the younger age of the 1965 and 1970 samples ought depress the mean performances of pupils, particularly in I.Q. and to some extent on the other tests. On the other hand, children repeating a grade may be expected to yield below average scores even in their second year in that grade.

With respect to the validity of the Australian Council for Educational Research norms, it should be noted that as long as analysis and interpretation is based upon raw scores then the year of original or subsequent test norming becomes irrelevant. The issue of changing emphasis in curricula is a more pertinent issue though does not necessarily destroy the usefulness of comparisons between test performances *per se*.

The differences in test administration time is a more significant factor. Certainly, the 1955 scores obtained in August, would be significantly depressed estimates of the November performance of the same sample of children. The 1970 survey was conducted on the last week of October.

RESULTS

Year Group Comparisons: Grade 7

Because of the emphasis on acceleration and retardation in the 1960 survey, the overall measure of central tendency was the median, rather than the mean.

The 1960 raw score data was not available for re-analysis, so that the only measure of central tendency from which comparisons of all four year groups might be made are the medians.

Strong parametric tests of differences between medians have not been developed. Hence, though the median scores for each year group may be calculated, statistical tests of differences such as 't' or 'F' are not appropriate.

The median scores on all major tests are shown in Table 5.

TABLE 5 : MEDIAN SCORES FOR 1955, 1960,* 1965, 1970 GRADE 7 SAMPLES

Test	1955	1960	1965	1970
Intelligence Test	97.8	102	101.0	104.9
Word Knowledge	42.6	48	40.8	39.8
Reading for Meaning	33.5	38	36.18	36.3
Addition	42.6	47	44.7	43.2
Subtraction	47.0	50	49.9	47.2
Multiplication	31.4	34	33.6	32.0
Division	33.3	36	33.3	30.0

The superior performance of the 1960 year group in most subjects is quite noticeable. Since 1960 a decline in the arithmetic and word knowledge median scores is also evident. On the other hand, the intelligence test results show the opposite trend. Setting aside the 1960 peak, the reading comprehension gains from 1955 to 1965 appear to have been maintained.

* The source for the 1960 median results is the 1965 survey report. Presumably they have been estimated from the frequency distributions tabled in the 1960 report. The figures are consistent with the distributions shown in Table 8.

TABLE 6 : MEANS AND STANDARD DEVIATIONS FOR THE 1970, 1965 and 1955
GRADE 7 SAMPLES ($N_{1970} = 1031$, $N_{1965} = 1147$, $N_{1955} = 917$)

Test	1970		1965		1955		Significance of Difference Between Means
	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.	
Intelligence Test	104.64	14.24	102.21	12.61	96.26	12.42	$p < .05$
Word Knowledge (10 min)	42.04	14.02	42.96	13.87	44.18	14.51	$p < .05$
Reading for Meaning (Comprehension)	37.80	13.93	35.49	11.33	34.84	10.35	$p < .05$
Addition	44.17	6.96	44.80	5.79	43.62	5.55	$p < .05$
Subtraction	48.03	10.12	49.98	9.87	47.96	8.72	$p < .05$
Multiplication	31.89	5.96	33.22	4.86	32.55	4.94	$p < .05$
Division	31.28	9.25	33.39	8.31	34.37	9.14	$p < .05$
Spelling (Words)	36.52	11.17	43.27	8.14	-	-	$p < .05$
Spelling (Sentences)	26.77	13.06	28.14	11.00	-	-	$p < .05$

Table 6, predictably corroborates the pattern of results evident from an inspection of the median scores. For reasons already discussed the 1960 means could not be tabled. A univariate analysis of variance established that on each test the difference between the largest and smallest year-group mean was statistically significant.

Apart from the mean differences, the relative size of standard deviations is of interest. In every instance the standard deviations of the 1970 sample, irrespective of the size of means, is larger than the standard deviations of the 1965 and 1955 samples. The standard deviation is a measure of dispersion of scores. This means that in the 1970 sample pupil test scores were more widely spread about the mean.

In order to provide a more complete comparison of performance across all four year groups frequency distributions were computed according to the manner adopted in the 1960 survey. The achievement test scores were transformed into Grade Point Averages for this purpose while the intelligence test results were transformed into standardized I.Q. test scores.

TABLE 7 : DISTRIBUTION OF STANDARDISED INTELLIGENCE TEST SCORES FOR
1955, 1960, 1965 AND 1970 SAMPLES BY PERCENTAGE OF SAMPLE

Intelligence Test Standard Score Range	1955 %	1960 %	1965 %	1970 %
69 -	1.0	0.3	0.4	.01
70 - 89	22.4	19.5	15.5	11.4
90 - 109	58.0	58.1	53.4	52.5
110 - 129	18.3	21.6	29.7	34.1
130 +	0.3	0.6	1.1	1.6

TABLE 8 : DISTRIBUTION OF GRADE POINT AVERAGE SCORES FOR GRADE 7
SAMPLE BY PERCENTAGE OF SAMPLE
WORD KNOWLEDGE

Grade Norms	1955	1960	1965	1970
Less than 4.0	1.0	.8	.9	.3
4.0 - 4.9	3.3	2.2	2.9	2.3
5.0 - 5.9	11.7	8.4	9.8	15.9
6.0 - 6.9	24.1	19.3	29.6	29.14
7.0 - 7.9	23.0	13.6	22.5	25.44
More than 7.9	37.0	55.7	34.4	34.1

READING FOR MEANING

Grade Norms	1955	1960	1965	1970
Less than 4.0	.5	.2	2.0	1.1
4.0 - 4.9	4.4	4.2	4.9	4.6
5.0 - 5.9	11.6	8.2	7.5	11.0
6.0 - 6.9	21.5	14.3	17.3	16.0
7.0 - 7.9	24.0	20.8	22.7	19.5
More than 7.9	38.1	52.4	45.5	47.9

DIVISION

Grade Norms	1955	1960	1965	1970
Less than 4.0	3.4	4.1	2.3	4.1
4.0 - 4.9	4.3	5.1	4.5	9.5
5.0 - 5.9	24.7	18.2	23.1	32.07
6.0 - 6.9	20.9	18.9	24.4	21.73
7.0 - 7.9	23.2	20.92	24.8	18.13
More than 7.9	23.4	32.7	20.9	14.52

MULTIPLICATION

Grade Norms	1955	1960	1965	1970
Less than 4.0	2.0	1.5	1.4	2.8
4.0 - 4.9	4.5	4.8	3.8	7.89
5.0 - 5.9	33.8	24.3	20.0	40.0
6.0 - 6.9	24.4	18.3	17.7	18.2
7.0 - 7.9	19.7	24.3	25.2	24.17
More than 7.9	15.6	26.7	32.0	15.7

ADDITION

Grade Norms	1955	1960	1965	1970
Less than 4.0	.9	.4	.6	1.0
4.0 - 4.9	7.7	5.3	4.3	6.1
5.0 - 5.9	20.3	12.7	12.5	19.61
6.0 - 6.9	24.5	17.9	19.3	22.3
7.0 - 7.9	22.2	13.5	24.8	18.6
More than 7.9	24.4	50.2	38.5	32.29

SUBTRACTION

Grade Norms	1955	1960	1965	1970
Less than 4.0	1.1	.6	1.4	1.0
4.0 - 4.9	3.3	2.7	2.7	3.7
5.0 - 5.9	17.8	16.0	11.4	18.0
6.0 - 6.9	27.8	17.9	19.8	26.37
7.0 - 7.9	26.5	20.3	21.6	19.24
More than 7.9	23.6	42.4	43.1	31.71

Table 8 substantiates the picture produced by the tabulation of mean and median scores. The dispersion of scores about the average is as expected and there appear to be no remarkable discrepancies between year groups with the exception of the 1960 sample. A discussion of the extraordinarily high performance of that year group will be discussed at a later point.

A second feature of the results is the continuing gain in performance of pupils in the I.Q. test over the four testing occasions. The peak performance of the 1970 group is all the more surprising in view of the relatively superior performance of the 1960 sample on virtually every other test performance. The implications of this result will receive a fuller discussion elsewhere.

The implications of Table 8 will depend upon the orientation of the reader. Certainly, at least it reinforces the appreciation that in any typical Grade 7 class there could well be a number of children unable to perform the skills at a standard that could be generally expected of an "average" child in Grade 3.

Year Group Comparisons: Grade 5

The pattern of performance of the Grade 7 students was not entirely replicated at the Grade 5 level, though the trend of relatively lower performance by the 1970 year group was continued in all areas apart from the intelligence test. However, only in multiplication, division and spelling (words) were the differences statistically significant. The test of difference between means employed was a multivariate analysis of variance.

However owing to the limitations of the computer (insufficient core space) the sample size was reduced by half by the systematic random sampling of alternate punch cards. The sub-sample and total sample means and standard deviations are shown in Table 9. Again, as with the Grade 7 analysis, the larger standard deviations of the 1970 year group are a significant feature.

TABLE 9 : MEANS AND STANDARD DEVIATIONS FOR THE REDUCED 1970 AND 1965 GRADE 5 SAMPLES ($N_{1970} = 511$, $N_{1965} = 526$)

Test	1970		1965		Significance of Difference Between Means
	$\bar{X}$	S.D.	$\bar{X}$	S.D.	
Intelligence Test	109.20	14.33	106.11	14.01	N.S.
Word Knowledge (10 min)	25.96	12.46	26.50	13.4	N.S.
Word Knowledge (15 min)	32.30	16.14	32.16	15.05	N.S.
Reading for Meaning	24.25	12.43	25.06	10.46	N.S.
Addition	36.70	8.65	37.43	7.37	N.S.
Subtraction	37.54	10.94	36.52	10.78	N.S.
Multiplication	24.48	7.99	26.69	7.35	$P < .05$
Division	21.87	9.38	24.32	9.08	$P < .05$
Spelling (Words)	39.96	9.60	43.81	8.74	$P < .05$
Spelling (Sentences)	22.93	13.25	24.16	15.00	N.S.

A Multiple Discriminant Analysis of the Grade 7 Results

Multiple Discriminant Analysis is a statistical technique which seeks not only to see whether a number of groups differ significantly in terms of a set of variables, but also to gain some insight into the nature of the differences. With a relatively large number of variables (or tests in this instance) it became difficult to interpret differences between groups in each variable taken singularly. A simple inspection of the means and standard deviations would reveal the tests with the greatest mean differences and might therefore indicate the tests which characterise differences in performances between groups. However, because the tests are intercorrelated, have different means, standard deviations and potential ranges of scores, the inspection may provide only a superficial picture of group differences.

One alternative is to take a weighted sum of the set of test scores that will minimise the differences between the groups. The weights, by

which each original test score is multiplied, are called discriminant weights or coefficients and these (slightly modified) will be tabled below.

The equation,

$$Y_1 = a_1 X_1 + a_2 X_2 + a_3 X_3 \text{ ----- } a_n X_n$$

is called the discriminant function. Y_1 is a discriminant score, $a_1 \dots a_n$ are the discriminant weights, and $X_1 \dots X_n$ are the test scores (X_1 = I.Q., X_2 = Word Knowledge, etc.).

With three groups (such as the 1955, 1965 and 1970 year groups) two functions may usually be extracted. Thus one or two functions may succinctly describe the year group differences where seven tests results were initially used. This is one of the real values of multiple discriminant analysis. By examining the mean discriminant scores for each group (these are known as centroids) and by examining the relative weights assigned to the different variables a deeper understanding of the group differences may be gained.

In the analysis of the 1970, 1965 and 1955 year groups two statistically significant functions were extracted. The chi-square test of significance for discrimination by the first dimension yielded a value of 471.99, significant well beyond the .001 level, while the second dimension, with a chi square value of 74.08, was also significant beyond the .001 level. The first vector (or function) accounted for 87 per cent of the predictable variation between the three year groups, while the second function accounted for the remaining predictable variation.

Correlations between the generated discriminant scores and the original seven test scores are shown in Table 10.

It can be noted from this data that the I.Q., reading comprehension and division tests have the highest loadings on the first discriminant function and largely characterise the dimension measured by this function. The ability to successfully complete the I.Q. and reading comprehension tests, as distinct from the division test tend to determine a pupil's score on the first function.

TABLE 10: CORRELATIONS BETWEEN DISCRIMINANT SCORES AND TEST SCORES

Test	Discriminant Function 1	Discriminant Function 2
Intelligence Test	0.41	-0.00
Word Knowledge	-0.16	-0.00
Reading for Meaning	0.26	-0.26
Addition	0.10	0.44
Subtraction	0.01	0.64
Multiplication	-0.12	0.62
Division	-0.36	0.26

That is, a student with a high discriminant score on this function would have high I.Q. and reading comprehension scores and relatively low division scores. This function has tentatively been labelled "verbal acuity". The relatively high loadings of the arithmetic tests on the second dimension tends to identify it as an "arithmetic-computation" function.

TABLE 11: GROUP CENTROIDS FOR DISCRIMINANT FUNCTIONS 1 AND 2 BASED ON AN ANALYSIS OF 1955, 1965 AND 1970 GRADE 7 RESULTS

Year Group	Function 1 "Verbal Acuity"	Function 2 "Arithmetic-Computation"
1970	39.12	24.88
1965	36.46	26.65
1955	33.18	24.93

Table 11 shows the group centroids (or mean discriminant scores) for each year group. The first discriminant function separates the three groups. The 1970 year group is characterised by greater verbal (perhaps verbal-intellectual) skill. Obviously, this result is influenced by the relatively high scores of the 1970 year group on the I.Q. and reading comprehension tests. The second function, the arithmetical-computation factor, discriminates the 1965 year groups from both the 1955 and 1970 year groups by virtue of the relatively higher, overall performance of pupils on the computation tests.

Grade 5 and Grade 7 Comparisons

As identical reading comprehension, word knowledge and arithmetic tests were administered to both the Grade 5 and Grade 7, 1965 and 1970 samples, a further basis of comparison was available. The results are shown in Table

TABLE 12: COMPARISONS BETWEEN GRADE 5 AND GRADE 7 MEAN SCORES FOR 1965 AND 1970 YEAR GROUPS

		1965	1970
Word Knowledge:	Grade 5	25.97	25.86
	Grade 7	42.96	42.04
Reading Comprehension:	Grade 5	24.43	24.05
	Grade 7	35.49	37.80
Addition:	Grade 5	37.46	36.67
	Grade 7	44.80	44.17
Subtraction:	Grade 5	36.31	37.28
	Grade 7	49.98	48.03
Multiplication:	Grade 5	27.44	24.11
	Grade 7	33.22	31.89
Division:	Grade 5	24.26	21.42
	Grade 7	33.39	31.28

The statistics in Table 12, apart from allowing some direct comparison of Grade 7 and Grade 5 standards (as measured by the particular test administered), permit more accurate estimate of performance trends and interactions.

For example both the reading comprehension and word knowledge tests indicate substantial differences between Grade 5 and Grade 7 samples. This difference apart from confirming the fact that students acquire considerable skill and knowledge between the Grade 5 and Grade 7 level, may also indicate that for Grade 5 students the tests are too difficult. The mean Grade 5 word knowledge score was only one quarter of the total possible score.

A further use of the Table above is to determine whether the trends for 1965 and 1970 are consistent for Grade 5 and Grade 7. In technical terms this question may be framed by asking whether there is an interaction between grade and year of survey. Probably the most significant interaction trend concerns the reading comprehension results. The relative improvement in Grade 7 reading is not apparent at the Grade 5 level. There is no corresponding interaction of this type in the closely related word knowledge.

Explaining this result is a speculative matter. Of course there is the possibility that the relative gain is due to sampling error. Or, the effect may well be due to some factor (such as experience of Grade 7 teachers or a curriculum innovation) which has not been considered in the survey.

The 1960 Peak

The relatively outstanding performance of the 1960 sample deserves some discussion. Was that year the highwater mark of Grade 7 achievement in the basic skills tested? What alternative explanations exist? A first point of verification might be an examination of sampling procedures. Anomalies in the selection of pupils may have introduced bias.

The 1960 report does not elaborate the procedures adopted in drawing the sample. It simply states that "During the last week in October 1960, 1,000 children in Grade 5 and 1,000 children in Grade 7 classes throughout the State were given a battery of tests". (p.1) However official records indicate that based on location and class of school, a proportionate random sample was selected. In other words, sampling procedures were comparable to those adopted in the other surveys.

A second possible source of error could have been the marking and scaling procedures adopted for the 1960 survey. However, the arrangements adopted for the 1960 survey were comparable to those adopted for the other surveys. Lists with full instructions for administering and marking tests were sent to schools. The class teachers then administered the tests, marked them from keys provided, entered results on the scoring sheets provided and returned to Head Office all tests, both used and unused, instructions and marking keys. In the 1960 survey, for the first time automatic data processing procedures were utilised and basic analyses, following full specification, were contracted to I.B.M. Thus while complete standardisation of test administration procedures could not be totally ensured,

there appears no reason to suggest why any anomaly might have occurred in 1960 rather than in any other survey.

The relatively high average test scores of the 1960 subgroup must stand as reasonable estimates of that particular sample's performance. For several of the tests, notwithstanding that statistical tests are not possible, the differences between median scores for that year and the other survey years seem of sufficient magnitude to rule out the caprice of sampling error as an explanation of events.

The Intelligence Test Results

The mean intelligence test scores showed a consistent upward trend over all survey occasions at both the Grade 5 and Grade 7 level. On none of the skill tests was this pattern repeated. The simplistic conclusion that follows is that children are becoming progressively "brighter", though not necessarily more able to read and compute.

The 1965 report noted this trend and suggested that factors to be considered were the exclusion of "special schools" and the lapse of time since the norming of the Australian Council for Educational Research's Intermediate Intelligence Test in 1949. While a renorming would have adjusted subsequent survey averages closer to the theoretical "average 100" this event would not help explain why in this particular intelligence test progressively over each five year period children have successfully completed more test items.

For some reason children appear to becoming more adept at responding to the items that compose the Australian Council for Educational Research Intelligence Test. Whether this facility is attributable to more than test wiseness (practice and exposure to comparable types of tests or test items) or, whether the schools have been progressively incorporating into their curricula components that might form the basis of an intelligence test, is open to question. An inquiry following the 1970 test administration indicated that many children had completed a form of the Australian Council for Educational Research Intelligence Test early in the year under the direction of Departmental Guidance Officers. However, both hypotheses seem plausible.

A perplexing aspect of this trend, assuming that it represents a real gain in ability to respond to intelligence tests, is that it has not been accompanied by similar patterns of performance on the skill tests. It has already been noted that most of the tests are at least moderately inter-correlated. For the Grade 7 sample the correlation coefficients ranged from 0.40 for addition to 0.72 for reading comprehension. This circumstance makes explanation more difficult. This phenomenon seems worthy of further investigation.

The Relationship Between School and Other Factors to School Achievement Need for Caution

It has already been discussed in this Report that the survey is not an appropriate means of determining the influence (or causal relationship) between school organisational factors and school achievement. This view can be easily demonstrated in Table 13: Mean Grade 7 Achievement Test Differences According to Certain School and Other Factors. Results from the survey in 1965 and 1970 have consistently favoured class sizes of forty or more pupils. The 1960 survey found that the highest levels of achievement occurred in classes of *fifty* or more children! Does this mean that the Education Department should rush towards *increasing* class size? Obviously not. There is no substantial evidence that class size, of itself, contributes to higher or lower school achievement. It could well be that more skilled teachers are found in larger classes, a deliberate strategy effected by the school principal. This hypothesis, like many others that could be advanced is largely speculative. The survey at least cannot substantiate or dismiss these explanations. In a similar manner there is little to be gained in attributing causal influences to the relationships that occur, say between location of school and achievement. This is not to say that the results in Table 13 are necessarily of no use. As *description* of achievement according to the various categories they are a valid record. Yet even as description the usefulness must be qualified. For example what is served by knowing that pupils in metropolitan schools did "better" than children in country schools at multiplication though worse at division. It seems absurd to think that there is some underlying quality of country schooling which inhibits multiplication skills yet facilitates division skills. Presumably, differences of this type are largely due to sampling error.

TABLE 13: MEAN GRADE 7 ACHIEVEMENT TEST DIFFERENCES ACCORDING TO CERTAIN SCHOOL AND OTHER FACTORS¹

Test	Sex of Pupil (+ favours boys)			Location of School (+ favours Metro.)			Class Size ² (+ favours larger)		Class Grouping ³ (+ favours grouped)		Streaming (+ favours streaming)		Class of School (+ favours Class 1)		
	55	65	70	55	65	70	65	70	65	70	65	70	55	65	70
Intelligence Test	+0.26	-0.39	+0.51	+3.22	+3.21	+1.37	+2.79	+5.71	-1.42	-1.44	-0.12	-1.31	-1.95	-1.07	-2.44
Word Knowledge	+1.67	-1.45	-1.40	+3.15	+3.19	-0.51	+2.95	+6.80	-2.35	-1.75	+0.75	-1.04	-4.19	-1.71	-1.26
Reading for Meaning	-0.46	-1.75	-2.25	+2.97	+0.61	-0.85	+1.19	+7.34	-0.44	-2.59	-1.18	+1.90	-3.43	-1.91	+0.65
Addition	-1.38	-1.13	-1.25	-0.91	+2.78	+1.15	+0.44	+1.59	-0.56	-2.59	+1.69	+0.50	+0.50	+1.58	-0.88
Subtraction	-1.17	-0.34	+0.05	+0.51	+3.17	+1.32	+1.32	+0.73	-1.81	+0.12	+2.63	+0.88	-1.70	+1.51	-1.64
Multiplication	-0.27	-0.33	-0.30	+0.99	+1.94	+0.54	+0.90	+1.86	-1.04	-0.61	+1.68	+0.05	-1.12	+1.71	-1.41
Division	-1.12	-0.84	-1.13	+2.30	+2.97	-0.06	+2.18	+2.74	-1.44	-1.75	+2.40	-0.44	-3.38	+0.85	-1.81
Spelling (Words)	N.A.	-2.39	-3.43		+3.40	+2.13	+1.96	+3.71	-2.34	-1.67	+2.92	-1.53	N.A.	-1.23	-2.33
Spelling (Sentences)	N.A.	-2.47			+3.90	+2.88	+2.04	+5.01	-1.80	-1.93	+2.78	-2.12	N.A.	+0.35	-2.10

¹ Differences are expressed in terms of raw score units.

² The mean differences reported in this column are drawn from comparisons between class size categories "31-40" and "40 or more pupils", where sub-sample sizes were in 1965, 274 and 722 pupils respectively and in 1970, 526 and 439 pupils respectively. These two categories contained nearly 90 per cent of the total sample of pupils.

³ The mean differences reported in this column are drawn from comparisons between pupils instructed in either single grade classrooms or classrooms which contain pupils in two grades. Pupils in classrooms containing more than two grades contributed to less than five per cent of the total sample. In the 1965 sample 871 pupils were in single grade classrooms and 187 in rooms containing two grades. In the 1970 sample 716 pupils were in single grade classrooms and 201 pupils in rooms containing two grades.

Sex Differences

The superior performance of girls in certain tests of educational achievement is well noted in the literature. A relatively faster rate of cognitive and emotional development is said to be one factor accounting for this phenomenon. Table 13 shows on most tests marginal advantages in favour of girls. However, it should be noted that many of the differences in mean performance between the sexes are very slight and could be due to sampling error. The occasional superior performance of boys, for example in word knowledge in 1955 and subtraction in 1970, suggests the need for some qualification of clear cut statements of expectation of superior performance on the part of girls.

Location of School

In 1965 the pattern of survey results showed a consistent advantage favouring the metropolitan schools. In 1970 the situation was less clear with most differences, excepting the spelling test results, of a very small magnitude. The position is now such that, in terms of the test battery employed in the survey, it is not possible to make unqualified statements regarding superior metropolitan or country performance.

Class Size

In order to examine the relationship between class size and achievement five categories were arbitrarily decided: 0-10, 11-20, 21-30, 31-40 and more than 40. The large majority of the sample for 1965 and 1970 (comparable data from 1955 and 1960 were not available) were accommodated in categories of "31-40" and "more than 40". Consequently only these two categories were utilised in subsequent analyses. As Table 13 indicates, without exception differences between mean performances of children in these categories favoured children in the larger class size. In the 1970 sample these differences are of considerable magnitude, particularly in the tests containing a "verbal" component.

The 1960 report discussed a similar trend. It noted that "The best performance in both grades (Grade 5 and Grade 7) were obtained by children in classes of 50 or over. On the composite scores in both grades, this group had less retardation and more accelerated children than any other group". (p.13)

Class Grouping

A grouped class in this study refers to a multi-grade class; a class where a single teacher is responsible for the supervision of two groups (or more) of students each largely and cohesively pursuing separate objectives and curricula. The situation occurs in those situations where because of insufficient numbers to justify the formation of separate classes a single teacher is given instructional responsibility for more than one grade.

It should be noted that class grouping, along with other variables employed in this survey, is a rather coarsely grained variable. It does not necessarily provide any picture of what instructional strategies are adopted in the classroom, and a great many different procedures may operate under a guise of 'mixed grades' or whatever nomenclature is used.

Relatively few students are enrolled in Government schools in this State which operate with a mixture of more than two grades in one classroom. Hence, the statistics derived for classes of three and more grades, because of the paucity of numbers, tend to be unreliable. In both the 1965 and 1970 surveys the class containing four grades tended to do better than the more grade-homogeneous groupings. When it is borne in mind that the sub-sample size for the 1965 and 1970 was 41 and 11 respectively then the need for caution becomes apparent. Consequently Table 13 contains statistics which relate to classrooms with single grade or collectivities of two grades. Results favour by small margins ungrouped classes with the single exception of the subtraction test.

Streaming

The division of children into classes according to general ability (streaming) has been a contentious issue for many years. Various educational advantages are attributed to instruction in either streamed or unstreamed class grouping. The literature tends to be equivocal. Unfortunately, Table 13 does not yield any more conclusive a picture. In both the 1965 and 1970 surveys there were superior performances for both the streamed and unstreamed situations. To add to the inconclusiveness on some tests in 1970 the 1965 pattern of performance was reversed.

It does not seem possible to make generalised statements favouring streamed or unstreamed conditions. Furthermore to conclude that streaming favours the development of subtraction computation skills yet not multiplication skills seems a fatuous statement.

Class of School

For administrative purposes schools are divided by the Education Department into various categories. Class IA schools are the largest with enrolments of approximately 700 children. Class IV schools are the smallest and are usually remote, one teacher schools with enrolments of perhaps 20 children. Thus in some respects "class of school" is a proxy variable for "school size".

In Table 13 comparisons are shown in terms of differences between performances of pupils in Class IA and Class II schools. Thus the comparisons are between children attending the largest and the middle range of primary schools. To have utilised Class IV instead of Class II schools in the comparison would have in essence repeated the metropolitan-country comparisons.

The results are not clear cut though a clear majority of comparisons favour the children attending the smaller Class II schools.

CONCLUSIONS

The 1955 Survey

The 1955 survey concludes:

1. The main purpose of the survey was to enquire into the amount of retardation. This was seen to be very slight in reading, whereas there appears to be a definite retardation in arithmetic.

The conclusion continues:

It seems possible that there might be some connection between the methods of instruction used and the amount of retardation. For some time reading has been approached from the point of individual differences. Children are heard read individually and generally are divided into three or more reading groups in primary schools, to the extent that this is standard practice. This is not so to the same extent for arithmetic. If individual differences can be cohered for arithmetic by some form of group instruction as is done by reading it should be possible to accomplish this within the present organisation with classes existing at present.

2. Because of the nature of the tests involving number combinations and tables, the fundamental processes, it would appear that these need constant repetition and revision up to Standard VI in arithmetic.

The report concludes on an optimistic note:

3. Care must be taken to see that the satisfactory standard of reading is maintained and that the small percentage of retarded readers is eliminated.

In terms of conclusion three, it could be argued that by 1970 the schools were meeting with some success. The number of children reading below their grade average had been reduced (though not eliminated) and the overall, standard appeared to have risen considerably.

The situation in arithmetic was more confused. Gains had been achieved in addition and subtraction while losses had occurred in multiplication and division.

The 1960 Survey

Performance of children in the 1960 survey was summarised as follows:

The results seem to indicate that a large reduction of retardation in reading has occurred over the five years between 1955 and 1960, and an increase in those children who were considered accelerated has also become obvious. Arithmetic shows some improvement but not to the same marked degree as in reading (p.11).

The improvement in reading was explained in these terms:

It seems possible that curriculum requirements and methods of instruction used might bear some relation to the amount of retardation. For some time teachers and research workers have paid a lot of attention to catering for individual differences in reading. Over the last five years individualized reading schemes, the recognition of the necessity to give children readers designed to cater for their individual abilities, reading readiness programmes and a greater effort on giving children instruction in specified reading skills have all helped to produce these results (p.12).

As in 1955 arithmetic results were categorised as "poor" even though some overall improvement had been noted.

Although improvement in arithmetic has taken place, this has not been due to any organized effort to produce these results.

In arithmetic, more material is needed to give children the opportunity to gain a knowledge of number by being able to understand it, rather than to develop it through rote drills and abstract experience. It has been said that the fundamental processes are of utmost importance in being able to manipulate numbers. If this is so, the poor results in arithmetic as measured by the tests used in the survey, will only be obviated by an effort to produce an approach to number based on understanding and a recognition that individuals should be able to progress at their own rate to the limit of their capacity. The Cuisenaire method of teaching number may be a solution to some of these problems (p.12).

The 1965 Survey

The report of the 1965 survey is confined largely to a description of results and contains little, if any judgmental observation regarding standards of reading or arithmetic.

The emphasis of the 1965 survey was primarily to determine the influence of selected factors such as location and size of school, school organisation, class size, bus travel and sex of pupil upon pupil achievement.

The report offered no recommendations for further action.

The 1970 Survey

1. Large scale longitudinal studies, such as the Survey of Primary School Achievement inherit particular problems of sampling and test reliability and validity. These are crucial aspects of any empirical investigation of this type for unless the sample is demonstrably representative of the population and the tests are of proven adequacy then any conclusions derived from the survey must be of doubtful value.

Of these two major considerations, sampling and instrumentation, the former is the more satisfactory in its present survey format. This is not to say that there could not afford to be modifications and refinements. For example, the precision of the sampling would be increased if, assuming that the present ceiling of approximately 1,000 pupils must remain, a larger number of schools were selected and some proportion randomly chosen from within each class or grade cohort.

Of more significance, is the nature of the instrumentation, that is, the test battery used to monitor standards.

The particular achievement tests appear to be pitched at an inappropriate difficulty level. They do not adequately discriminate. They test only a very narrow perspective of primary school achievement. It would appear that more appropriate tests could be adapted or constructed to either replace or supplement the existing test battery in order to give the survey the depth and meaning that are warranted by the last decade of curriculum innovation. Without this modification, a future survey does not seem warranted.

It cannot be emphasised too strongly that without a proper understanding of the nature of the tests used in the survey, statements about decline or improvement in "standards" are likely to be highly misleading. Many teachers in 1970 would dismiss the attributes being measured by several of the tests, particularly those relating to arithmetic computation, as bordering on the trivial in comparison to the total curriculum of the school

2. At the Grade 7 level the overall trends in primary school achievement over the past fifteen years appears to have been that performance in most of the skills listed reached a peak in 1960 and deteriorated to some extent after that year. 1965 and 1970 comparisons revealed the 1965 year group to

be at an advantage (though in some subjects this advantage was nearly negligible) in all subjects except reading comprehension.

It is worth noting in an attempt to provide balance that the 1955 report concluded that the standard of reading was satisfactory and should be maintained. In fact in 1970 performance on the reading comprehension test had risen by a statistically significant amount. There is obviously no absolute standard of performance.

At the Grade 5 level, the discrepancies between the 1965 and 1970 year groups favoured the 1965 students.

In 1970 there was a greater dispersion of test scores than in any other survey year.

Interpreting these results is largely speculative owing to the inadequacies of the survey as a research method. It is unlikely that any single factor is responsible for the differences between the year groups. One tentative explanation put forward is that the particular skills measured by the survey test battery have received less emphasis in 1970 owing to substantial curriculum change. Teachers in 1970 may not have stressed these skills to the same extent as their 1960 counterparts.

3. In general, it is true of the 1970 year group that according to the test performances:

- i. Girls out performed boys on a majority of tests.
- ii. There did not exist a clear and consistent pattern of results favouring either metropolitan or country pupils.
- iii. Pupils in the larger classes gained on average higher scores than pupils in smaller classes.
- iv. Pupils in single graded classes out performed pupils in classes with mixed grades.
- v. Pupils in classes streamed according to general ability did not consistently gain on average higher test scores than pupils in unstreamed classes.
- vi. Pupils in Class II primary schools tended to gain higher average scores on most tests than children in Class IA primary schools.

4. Consideration must be given to the prospect of a 1975 survey. In view of the limitations regarding the current test battery an assessment must be made whether the likely outcomes of the survey warrant the investment of resources. To replace components of the test battery destroys the longitudinal component of the project. Hence a question must be posed regarding the value of continuing to administer the existing tests or whether a fresh start to monitoring standards is required.

REFERENCES

Crane, A.R. An Historical and Critical Account of the Accomplishment Quotient Idea. *The British Journal of Educational Psychology*. 1959, 29, 252-259.

Cronbach, L.J. and Warrington, W.G. Time-limit Tests: Estimating Their Reliability and Degree of Speeding. *Psychometrika*. 1951, 16, 167-188.

Flanagan, J.C. Units, Scores and Norms in E.F. Lindquist (Ed.). Educational Measurement American Council on Education. Washington, 1951.

Vernon, P.E. Dullness and Its Causes. *Times Educational Supplement*. 26th October, 1956, p.1277.

APPENDIX 1: CLASSIFICATION BY SCHOOL OF 1960, 1965 AND 1970 GRADE 7
SAMPLES

Classification	Total Population			Sample		
	%			%		
	1960	1965	1970	1960	1965	1970
Junior High School	12.0	10.1	-	12.0	-	-
Class IA	-	22.3	25.9	-	25.33	38.84
Class I	64.0	44.2	64.0	64.0	49.26	40.02
Class II	8.0	11.8	8.0	8.0	12.36	8.67
Class III	10.0	9.7	10.0	10.0	11.84	10.03
Class IV	4.0	1.2	4.0	4.0	0.01	0.02
Mission		.04	.08	2.0	-	-
Correspondence	2.0	.02	.02	-	-	-
Total	100.0	100.0	100.0	100.0	100.0	100.0

**APPENDIX 2: TESTS ADMINISTERED TO GRADE 7 SAMPLE IN 1955, 1960, 1965
AND 1970 SURVEYS**

Test	Time Limit	No. of Items	Reliability [†]	1955	1960	1965	1970
<u>Intelligence</u>							
A.C.E.R. Intermediate D	30	80		✓	✓	✓	✓
<u>Reading</u>							
A.C.E.R. Reading for Meaning Form D	20	60		✓	✓	✓	✓
A.C.E.R. Reading for Meaning Form D	20	60				✓	
A.C.E.R. Word Knowledge Form C	10*	100		✓	✓	✓	✓
A.C.E.R. Word Knowledge Form D	10	100				✓	
A.C.E.R. Silent Reading Form C		60		✓			
A.C.E.R. Silent Reading Form D		60				✓	
<u>Spelling</u>							
W.A. Grade 7 Spelling Test		49	?		✓	✓	✓
A.C.E.R. Spelling Test Form C		102	0.94		✓	✓	✓
<u>Arithmetic</u>							
<u>A.C.E.R. Arithmetic Form C</u>							
Part 1 Addition	70	50	0.85	✓	✓	✓	✓
Part 2 Subtraction	75	55	0.79	✓	✓	✓	✓
Part 3 Multiplication	55	35	0.73	✓	✓	✓	✓
Part 4 Division	60	40	0.93	✓	✓	✓	✓
<u>Departmental Arithmetic</u>							
Part 1			?		✓		
Part 2			?		✓		
Terms and Relations		60	0.75	✓			

* This test was administered with 10 minute and 15 minute time limits in 1965 and 1970. The standard time (reported in manuals) was 10 minutes.

† Data from test manuals.

APPENDIX 3: SAMPLE PRACTICE ITEMS1. A.C.E.R. Intermediate Test D

1. FOOT is to MAN as HOOF is to (?)
 (1) dog (2) horse (3) lion (4) cat (5) bird ()
2. What is the next number in this series?
 5, 10, 15, 20, 25, ... ()

2. A.C.E.R. Word Knowledge

1. quick	1. soft	2. quiet	3. fast	4. run	4. sharp (3)
2. small	1. man	2. baby	3. big	4. little	5. ant ()

3. A.C.E.R. Reading for Meaning

The trees were losing their leaves of crimson, gold and brown. As they fell, Old Tom the gardener swept them on to burning heaps from which the blue smoke drifted lazily.

The best name for this story is

1. Spring
2. Painting
3. Autumn Leaves
4. Rubbish
5. New Brooms

4. A.C.E.R. Arithmetic Test

Addition: 7
 310
 88
 57

Subtraction: 718
 639

Multiplication: 386
 78

Division: $9 \overline{)600}$

44.

5. A.C.E.R. Spelling (Sentences)

(Children write complete sentences as dictated by a Supervisor)

The completion of the task/was greeted with shouts of approval.

6. Western Australian Education Department Spelling (Words)

SWOLLEN The boy's broken arm became swollen SWOLLEN.

APPENDIX 4: ESTIMATING THE INFLUENCE OF SPEEDEDNESS ON THE PERFORMANCE OF THE WORD KNOWLEDGE TEST

Rarely are pure speed or pure power tests ever used in practice. Most tests are partially speeded to some degree, depending on how much time is allocated to the test. Cronbach and Warrington (1951) have provided a means of estimating the effect of the speededness of a test, which may limit the amount of time pupils may spend on a test.

Results from the 1965 survey has been used for this purpose for in that year data were collected from pupils under speeded and non-speeded (or power) conditions.

Parallel forms of the word knowledge test (forms C and D) were administered to the sample. Each form was administered under speeded conditions (10 minutes) and relative power conditions (here 15 minutes was allowed, though more time may have been desirable).

The formula is:

$$T = 1 - \frac{r(S_c, P_d) \cdot r(S_d, P_c)}{r(S_c, G_d) \cdot r(P_c, P_d)}$$

where S = Speeded Conditions

P = Power Conditions

C & D = Test forms

r = Product Moment Correlation

T = True Score Variance Attributed to Speededness

The outcome of this investigation is that speededness appeared to play a relatively small part in determining the ordering of students on the word knowledge test, at least for 1965, though the estimates of T are on the conservative side. At the Grade 7 level only 4.32 per cent of the variation of the students' word knowledge test performance is attributable to the speededness factor while at the Grade 5 level only 0.1 per cent.

The example above does not provide any information regarding whether this effect operated differentially during the four surveys. It is interesting to note results in the Table below for the word knowledge test, which was administered under both speeded and unspeeded conditions.

SIGNIFICANCE OF DIFFERENCE BETWEEN 1965 AND 1970 MEANS OF WORD KNOWLEDGE
TEST ADMINISTERED UNDER SPEEDED AND UNSPEEDED CONDITIONS

	Conditions	1965 $\bar{X}$	1970 $\bar{X}$	Difference $P = .05$ $\bar{X}$
Grade 7	Speeded (10 min.)	42.96	42.04	Non-Significant
	Unspeeded (15 min.)	54.74	52.36	Significant
Grade 5	Speeded (10 min.)	26.50	25.96	Non-Significant
	Unspeeded (15 min.)	32.16	32.30	Significant

While the differences in time limit did not appear to affect the relative standing of the two year groups at the Grade 5 level, it did at the Grade 7 level where, with the additional time, the 1965 year group significantly outscored the 1970 group. This result does to some extent substantiate the hypothesis advanced earlier that the ability to compute under speeded conditions was a much more important aim prior to 1970 and may therefore partly explain relatively poor performance of the 1970 year group on various tests.